REQUIREMENTS ENGINEERING FOR TARGET DRIVEN DATA WAREHOUSES

SHORT SYNOPSIS

Submitted in partial fulfillment of the requirement of the degree of

DOCTOR OF PHILOSOPHY

to

Manav Rachna University

by

Hanu Bhardwaj
2K16-PFE-008

Under the supervision of

Prof. (Dr.) Jyoti Pruthi
(Supervisor)

Prof. (Dr.) Naveen Prakash
(Co-Supervisor)

Department of Computer Science & Technology
School of Engineering

MANAV RACHNA
Vidyanatariksha

MANAV RACHNA UNIVERSITY
Declared as State Private University vide Haryana Act 26 of 2014

Year 2018

MANAV RACHNA EDUCATIONAL INSTITUTIONS

MANAV RACHNA UNIVERSITY

Established, vide Haryana Act no.26 of 2014
(Formerly Manav Rachna College of Engineering,
NAAC Accredited 'A' Grade Institute)

DECLARATION BY THE SCHOLAR

I declare that the Synopsis entitled **"Requirements Engineering for Target Driven Data Warehouses"**, prepared under the supervision of Dr. Jyoti Pruthi and Prof. Naveen Prakash, is being submitted for the approval of Departmental Research Committee (DRC).

19/12/17

Signature of Supervisor

Signature of Co-Supervisor

Signature of Scholar

INDEX

S. No.	Description	Page Nos.
1.	Abstract	1
2.	Identification of Research Problem	1-2
3.	Statement of Problem	2-3
3.	Literature Review	3-4
4.	Objectives of the Study	4
5.	Methodology To Be Adopted	4-5
6.	Proposed Time Frame	6
7.	Justification of the Study	6
8.	References	6-7

1. ABSTRACT

Requirements engineering for data warehouses has gained importance over time. We focus on former stage for data warehouse requirements engineering, i.e. an 'early information' part wherein relevant information for decision making is discovered. Early information data warehouse requirements engineering starts with targets defined as pairs of the form <A, I> wherein A is an organizational aspect and I is a business indicators' set. Targets are organized in a target hierarchy. This hierarchy is a complete specification of what is to be achieved by a top level target. We associate targets with choice sets so that alternative ways of target achievement can be represented. These alternatives form their own hierarchy. Finally, information relevant to selection of each alternative is discovered through Ends, Means, Key Success Factor, and Outcome Feedback analysis techniques. These techniques determine early information that is to be subsequently to be processed in the 'late information' requirements engineering stage. Appropriate techniques need to be identified for the elicitation of targets, sub targets, choice sets etc.

2. IDENTIFICATION OF RESEARCH PROBLEM

Requirements engineering in Software Engineering is the discipline concerned with understanding and documenting software requirements. Requirements need to be gathered through an elicitation process, before they can be analyzed or modeled. Thus the requirements engineering process starts with requirements elicitation, which is followed by requirements analysis and specification. Elicitation of Requirements is a process through which clients and developers discuss, review and understand the needs of the clients and the constraints on the software. The common techniques used for requirements elicitation, as listed in BABoK (Business Analyst Body of Knowledge), are Interviews, Questionnaires, Brainstorming, Prototyping/Modeling, Meeting focus groups, Inspecting documents, Observations etc.

Data Warehouses can be applicable in different domains for assisting management in planning business processes, evaluation of performance of business processes, and support in making decisions in business processes. DW's main objective is to help in making decisions after analysing distributed and heterogeneous information.

There is a significant difference in analyzing requirements for data warehouse systems as compared to analyzing requirements for traditional information systems. Difference lies in the nature of the problem addressed by the two kinds of systems itself. Traditional systems are 'transaction' or 'function' oriented and naturally, therefore, there is a need to discover and specify functions for the system to-be. On the other hand, in data warehouses, no 'functions' are implemented. Rather, a collection of data that shall support the decision making capability is to be identified. Therefore, the requirements engineering problem here is find a way to identify this information and to structure it appropriately.

In its early years, data warehouse development was centered on the design problem. Approach followed by Bill Inmon for data warehouse is top-down which starts with design of normalized data model. This is followed by the design of dimensional data marts centered around specific business processes or departments. Top-down approach has the disadvantage of representing a large project with a wide scope. Also it involves a significant up-front cost for implementation of data warehouse, and there is a substantial time lag between start of project till the point where initial benefits can be experienced by the end users. This top-down approach is unresponsive and inflexible to varying needs of the departments at the time of implementation.

Bottom up approach for designing a data warehouse has been given by Ralph Kimball. In this approach, creation of data marts is done first for providing reports and analysis for specific business processes, although bottom-up approach used in Kimball methodology is the outcome of business oriented top down analysis of the relevant business processes to be considered for modeling.

3. STATEMENT OF PROBLEM

We have observed that the existing problem of requirements engineering in data warehouse is to define the information of a data warehouse to support the decision making task at hand.

So we propose to partition this task into two parts:

a) 'early information', wherein relevant information for decision making is discovered.

b) ‘late’, wherein the structuring of the information discovered is done as facts and dimensions.

Our focus will be on the former, i.e. devise a way to elicit information that will go into the data warehouse for supporting in making decisions.

4. LITERATURE REVIEW

Requirements Engineering for Data Warehouses is gaining significance over the last years. For data warehouse development, three life cycles have been referred [Pra08] and [Pra09], which are driven by Database Schema, Entity relationship diagram and Goals.

The life cycle discussed by Golfarelli in [Gol99] starts with development of database schema. It is based on the assumption that information contents of the data warehouse system are most likely defined by the contents of the data in operational databases. The idea behind designing the data warehouse follows the approach of exploring each relation of the database manually and makes a decision to adopt the same as fact/dimension in data warehouse. Relational databases do not support decision making because they are designed for routine transactions. Also only the data existing in the database is considered, ignoring the data from external sources, which results in limited decision making capability of the data warehouse constructed.

Now consider the life cycle which is based on Entity Relationship Diagram as proposed by Hüseman in [Hüs00]. The source of constructing ERD is either the relational databases or the operational systems. The limitation of considering only the ERD information and ignoring the information from external or internal sources is faced in this approach. And moreover, this approach is not appropriate for handling complicated and random queries, which is the lifeline of data warehouses. ERD results in extremely normalized relations, which does not lead to a good data warehouse design which can handle intuitive queries. Kimball also stated "Entity relation models cannot be used as the basis for enterprise data warehouses", which discourages the use of ERD for data warehouse development.

Boehnlein, Golfarelli in [Boe99,Gol99] discuss the third approach of designing DW that is based on goals of the organization and stakeholders involved. Horkoff in [Hor12] also states that "goal modeling is not yet widely used in practice", although it has been applicable in case studies. Rolland in [Rol03] discusses about the factors which stand in the way of adopting techniques based on goals : "(a) Domain experts find the notion of a goal to be fuzzy which makes it difficult to deal with it, and (b) Goal reduction is not a straight forward process". Thus the similar issues are faced in requirement engineering for data warehouses as well.

5. Objective

Our proposal here is directly related to determining the information that shall be stored in the data warehouse. The outcome of our requirements engineering step is not the multi dimensional schema but early information that is unstructured. It identifies at a top level the information that is needed and that is to be structured in subsequent stages.

6. Methodology

Our proposal for early information requirements engineering starts off by determining targets that organizations want to achieve. We define target as a pair, <A, I> where A is an organizational aspect and I is a set of business indicators. An aspect represents a work area, work unit, service or quality in an organization. Business indicators are measures/metrics of aspect performance.

The notion of target hierarchy is proposed which involves target reduction. This can be done by following either the aspect driven or indicator driven approach. In the former, sub aspects of a top level target are determined and their business indicators are found to form sub targets. In the latter, sub business indicators that go into computing the top level business indicators are determined, their aspects found and sub targets formulated. Thus, we have a relatively clear basis for guiding target reduction as compared to goal reduction. We refer the target hierarchy as relevance hierarchy also.

Whereas target hierarchies identify what the organization wants to achieve, decision making is to be done to indeed achieve targets. Now we consider the second step of our 'early information'

requirements engineering stage. In this, we associate choice sets with targets. We define a choice set as a pair <CSO, O> where CSO is the choice set objective and O is the set of alternatives for meeting the CSO. And similar to target reduction, we have choice set reduction. Each sub choice set is associated with the target that it affects. This hierarchy is referred as fulfillment hierarchy.

At this moment, the full decisional capability required for target achievement is known. Now, in the last stage of early information requirements engineering, we discover the information required to make the most suitable selections from choice sets.

We observe that Indicators are the 'sticking glue' between the relevance and fulfillment dimensions, (i) they identify 'what is to be achieved' in the relevance hierarchy and (ii) can be used as a basis for defining 'how it is to be achieved' in the fulfillment dimension.

Once choice sets have been discovered in the fulfillment dimension, the issue is of determining the relevant information for alternative selection. This information is discovered through Ends, Means, Outcome Feedback, and Key Success Factor analysis techniques. The discovered information is 'early', abstract, fuzzy, and unstructured. It is to be converted to facts and dimensions in subsequent stages using the techniques spelled out in [Pra08].

Thus we trace requirements engineering process and the various steps comprising the process are as follows:

1. Determine top level aspects.
2. Determine business indicators.
3. Formulate top level targets.
4. Decompose Targets.
5. Identify Choice Set Objective could occur at varying times.
6. Build Achieves Hierarchy.
7. Identify Information.

7. PROPOSED TIME FRAME

The proposed research work will be completed in the following phases:

Phase I – Literature Survey of the broad area of research.

Phase II – Literature survey of the latest trends available in requirements engineering in data warehouses

Phase III – Design and development of a new better way of requirements gathering based on business indicator.

Phase IV – Analysis of results and conclusion.

Phase V – Thesis.

It is envisaged that the research work will progress as per the detailed plan.

8. JUSTIFICATION OF THE STUDY

The outcome of this dissertation entitled "REQUIREMENTS ENGINEERING FOR TARGET DRIVEN DATA WAREHOUSES" shall be to propose a reliable method of gathering requirements for a data warehouse. It is hoped that the research investigations reported and suggestions made thereof would be very useful to the industry engaged in building data warehouses.

9. References

[Bar11] Barone D., Jiang L., Amyot D., Mylopoulos J.: Composite Indicators for Business Intelligence, ER 2011, LNCS Vol. 6998, pp. 448–458 (2011)

[Boe99] Boehnlein, M., Ulbrich vom Ende, A.: "Deriving initial Data Warehouse Structures from the Conceptual Data Models of the Underlying Operational Information Systems" Proc. Of Workshop on Data Warehousing and OLAP (DOPLAP), Kansas City, MO, USA,1999

[Bon01] Bonifati A. ,. attaneo F.., CeriS., A. Fuggetta, and S. Paraboschi. Designing data marts for data warehouses. ACMTOSEM, 10(4):452–483, 2001.

[Fra08] Frank U., Heise D., Kattenstroth H., Schauer H., Designing and Utilizing Indicator Systems Within Enterprise Models – Outline of a Method, in MobIS, 89-105 (2008)

[Gog93] Goguen, Joseph A., and Charlotte Linde, "Techniques for Requirements Elicitation," *Proceedings of the International Symposium on Requirements Engineering (Colorado Springs, CO, April 18-22)* IEEE Computer Society Press, Los Alamitos, CA, 1993.

[Gol98] Matteo Golfarelli, Dario Maio, Stefano Rizzi: The Dimensional Fact Model: A Conceptual Model for Data Warehouses. Int. J. Cooperative Inf. Syst. 7(2-3): 215-247 (1998)

[Hor12] Horkoff J., and Yu E, Comparison and Evaluation of Goal-oriented Satisfaction Analysis Techniques, Requirement Engineering Journal, Springer, 2012

[Hüs00] Hüsemann, B., Lechtenbörger, J., Vossen, G.: Conceptual Data Warehouse Design. In: Proceedings of the International Workshop on Design and Management of Data Warehouses (DMDW 2000) Stockholm, Sweden, pp. 5–6 (2000)

[Kim96] Kimball R., The Data Warehouse toolkit, Wiley 1996.

[Kum10] Kumar M., Gosain A., Singh Y.: Stakeholders driven requirements engineering approach for data warehouse development, Journal of Information Processing Systems, Vol. 6, 3, pp. 385-402 (2010)

[Pra04] Prakash N, Singh Y, Gosain A., Informational Scenarios for Data Warehouse Requirements Specification, in Conceptual Modeling-ER'2004, Atzeni P., Chu W., Zhou S., Ling T K(eds.) LNCS 3288, Springer, 205-216, 2004

[Pra08] Prakash N., and Gosain A., An Approach to Engineering the Requirements of Data Warehouses, Requirements Engineering Journal, Springer, 13 (1), 49-72, 2008

[Pra09] Prakash N., Prakash D., and Sharma Y.K., Towards Better Fitting Data Warehouse Systems, A. Persson and J. Stirna (Eds.): PoEM 2009, LNBIP 39, pp. 130–144, 2009.

[Pra12] Prakash N., Bhardwaj H.: Early Information Requirements Engineering for Target Driven Data Warehouse Development, PoEM 2012, K.Sandkuhl, U. Seigerroth, and J. Stirna (Eds) LNBIP 134, Springer, pp. 188-202 (2012)

[Pra06] Prat N., Akoka J., Comyn-Attiau I.: A UML-based data warehouse design method, Decision Support Systems, Volume 42, Issue 3, pp. 1449–1473 (2006)

[Rol03] Rolland C., Reasoning with Goals to Engineer Requirements, Invited talk, Int. Conf. on Enterprise Information Systems, ICEIS 2003

[Sha10] Shanks G., Sharma R., Seddon P., Reynolds P., Business Analytics and Competitive Advantage: A Review and Research Agenda, in 21st Australasia Conference on Information Systems, pp. 1-11 (2010)

MANAV RACHNA
vidyanatariksha

Swami Vivekananda Library and Resource Centre
Manav Rachna University

SIMILARITY VERIFICATION CERTIFICATE

Title: Requirements Engineering for Target Driven Data Warehouses

Researcher: Hanu Bhardwaj

Registration Number: 2K16-PFE-008

Supervisor: Dr Jyoti Pruthi

Department: Computer Science & Technology Department

This is to certify that the above mentioned Short Synopsis/Thesis was scanned for similarity index through Turnitin Software on 17-Jan-2018 with submission ID 903535483 and the similarity index was found to be 3 %. The complete report running into 7 Pages, 2163 Word count with File Size 27.24 KB is being forwarded to the Research Supervisor/DRC/DRB for review.

Date: 17th Jan 2018

MANAV RACHNA UNIVERSITY
FARIDABAD

The complete report of the above Short Synopsis/Thesis has been reviewed and the similarity index is below the accepted norms as per the University Grants Commission (Promotion of Academic Integrity and Prevention of Plagiarism in Higher Educational Institutions) Regulations 2018. The Short Synopsis/Thesis may be considered for further processing

Hanu 17/01/18
Research Scholar

Jyoti 17/01/18
Research Supervisor

Hanu Bhardwaj Jan 17 2018

by Hanu Bhardwaj

Submission date: 17-Jan-2018 02:21PM (UTC+0530)
Submission ID: 903535483
File name: Synopsis_Hanu_17012018_check.docx (27.24K)
Word count: 2163
Character count: 12267

1. ABSTRACT

Requirements engineering for data warehouses has gained importance over time. We focus on former stage for data warehouse requirements engineering, i.e. an 'early information' part wherein relevant information for decision making is discovered. Early information data warehouse requirements engineering starts with targets defined as pairs of the form <A, I> wherein A is an organizational aspect and I is a business indicators' set. Targets are organized in a target hierarchy. This hierarchy is a complete specification of what is to be achieved by a top level target. We associate targets with choice sets so that alternative ways of target achievement can be represented. These alternatives form their own hierarchy. Finally, information relevant to selection of each alternative is discovered through Ends, Means, Key Success Factor, and Outcome Feedback analysis techniques. These techniques determine early information that is to be subsequently to be processed in the 'late information' requirements engineering stage. Appropriate techniques need to be identified for the elicitation of targets, sub targets, choice sets etc.

2. IDENTIFICATION OF RESEARCH PROBLEM

Requirements engineering in Software Engineering is the discipline concerned with understanding and documenting software requirements. Requirements need to be gathered through an elicitation process, before they can be analyzed or modeled. Thus the requirements engineering process starts with requirements elicitation, which is followed by requirements analysis and specification. Elicitation of Requirements is a process through which clients and developers discuss, review and understand the needs of the clients and the constraints on the software. The common techniques used for requirements elicitation, as listed in BABoK (Business Analyst Body of Knowledge), are Interviews, Questionnaires, Brainstorming, Prototyping/Modeling, Meeting focus groups, Inspecting documents, Observations etc.

Data Warehouses can be applicable in different domains for assisting management in planning business processes, evaluation of performance of business processes, and support in making decisions in business processes. DW's main objective is to help in making decisions after analysing distributed and heterogeneous information.

Hanu Bhardwaj Jan 17 2018

ORIGINALITY REPORT

Excluding matches less than 14 words

3%	1%	2%	1%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to University of Houston System Student Paper	1%
2	Submitted to Institute of Management Technology Student Paper	1%
3	Naveen Prakash. "An approach to engineering the requirements of data warehouses", Requirements Engineering, 01/2008 Publication	1%
4	mediatum.ub.tum.de Internet Source	<1%
5	www.irma-international.org Internet Source	<1%

Exclude quotes On

Exclude bibliography On

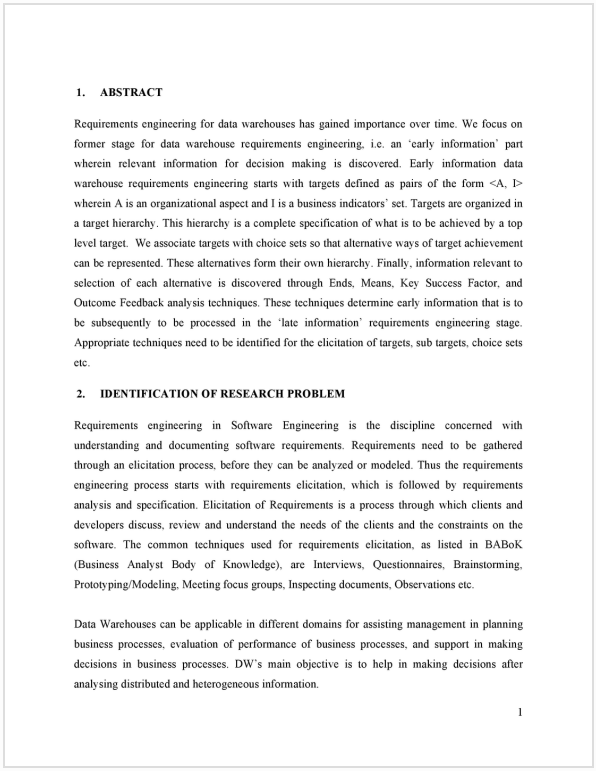
Exclude matches Off

Digital Receipt

This receipt acknowledges that Turnitin received your paper. Below you will find the receipt information regarding your submission.

The first page of your submissions is displayed below.

Submission author:	Hanu Bhardwaj
Assignment title:	Check Thesis
Submission title:	Hanu Bhardwaj Jan 17 2018
File name:	Synopsis_Hanu_17012018_check....
File size:	27.24K
Page count:	7
Word count:	2,163
Character count:	12,267
Submission date:	17-Jan-2018 02:21PM (UTC+0530)
Submission ID:	903535483



1. ABSTRACT

Requirements engineering for data warehouses has gained importance over time. We focus on former stage for data warehouse requirements engineering, i.e. an 'early information' part wherein relevant information for decision making is discovered. Early information data warehouse requirements engineering starts with targets defined as pairs of the form <A, I> wherein A is an organizational aspect and I is a business indicators' set. Targets are organized in a target hierarchy. This hierarchy is a complete specification of what is to be achieved by a top level target. We associate targets with choice sets so that alternative ways of target achievement can be represented. These alternatives form their own hierarchy. Finally, information relevant to selection of each alternative is discovered through Ends, Means, Key Success Factor, and Outcome Feedback analysis techniques. These techniques determine early information that is to be subsequently to be processed in the 'late information' requirements engineering stage. Appropriate techniques need to be identified for the elicitation of targets, sub targets, choice sets etc.

2. IDENTIFICATION OF RESEARCH PROBLEM

Requirements engineering in Software Engineering is the discipline concerned with understanding and documenting software requirements. Requirements need to be gathered through an elicitation process, before they can be analyzed or modeled. Thus the requirements engineering process starts with requirements elicitation, which is followed by requirements analysis and specification. Elicitation of Requirements is a process through which clients and developers discuss, review and understand the needs of the clients and the constraints on the software. The common techniques used for requirements elicitation, as listed in BABoK (Business Analyst Body of Knowledge), are Interviews, Questionnaires, Brainstorming, Prototyping/Modeling, Meeting focus groups, Inspecting documents, Observations etc.

Data Warehouses can be applicable in different domains for assisting management in planning business processes, evaluation of performance of business processes, and support in making decisions in business processes. DW's main objective is to help in making decisions after analysing distributed and heterogeneous information.

1

Copyright 2018 Turnitin. All rights reserved.